NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

REQUIREMENTS ANALYSIS FOR THE DEVELOPMENT OF DIGITAL LIBRARY FOR THE DOD INFORMATION OPERATIONS CENTER FOR EXCELLENCE (IOCFE)

by

Trisha Francis

September 2006

Thesis Advisor: Steven Iatrou
Second Reader Karl Pfeiffer

Approved for public release; distribution is unlimited

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704-0188
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington DC 20503.			
1. AGENCY USE ONLY ***(Leave blank)***	**2. REPORT DATE** September 2006	**3. REPORT TYPE AND DATES COVERED** Master's Thesis	
4. TITLE AND SUBTITLE: Requirements Analysis for the Development of Digital Library for the Dod Information Operations Center for Excellence (IOCFE)			**5. FUNDING NUMBERS**
6. AUTHOR(S) LT Trisha Francis			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000			**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES) N/A			**10. SPONSORING/MONITORING AGENCY REPORT NUMBER**
11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.			
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution is unlimited			**12b. DISTRIBUTION CODE** A

13. ABSTRACT (maximum 200 words)

In a memo from Paul Wolfowitz, Deputy Secretary of Defense, "The Naval Postgraduate School (NPS) is hereby designated the DoD Information Operations Center for Excellence. In that capacity, NPS shall facilitate development of Information Operations as a core military competency and innovation." Commander, US Strategic Command (USSTRATCOM) will serve as Operational Sponsor for the Center on behalf of the Combatant Commands. The Secretary of the Navy and Commander USSTRATCOM will develop a charter for the Center on Wolfowitz's approval, in coordination with the Under Secretaries of Defense for Policy and Intelligence, the Chairman of the Joint Chiefs of Staff, and other DoD officials as appropriate. The charter will address oversight and activities of the Center, including graduate education, research, research opportunities, and transformation. As a tool to enhance the IOCFE USSTRATCOM is looking into the development of a digital library which will specifically provide resources for the Information Operations Community. This thesis conducts a preliminary requirements analysis for the development of a digital library. Successful development of this digital library is expected to effectively enhance the operational areas of Information Operations and Information Warfare within the Department of Defense

14. SUBJECT TERMS Information Superiority, Electronic Warfare, Information Systems Technology, Sensors, Electronics & Electronic Warfare, Common/Consistent Tactical Picture, Knowledge Superiority and Assurance, Automation, Web-Enabled			**15. NUMBER OF PAGES** 59
			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT Unclassified	**18. SECURITY CLASSIFICATION OF THIS PAGE** Unclassified	**19. SECURITY CLASSIFICATION OF ABSTRACT** Unclassified	**20. LIMITATION OF ABSTRACT** UL

NSN 7540-01-280-5500 Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release; distribution unlimited

REQUIREMENTS ANALYSIS FOR THE DEVELOPMENT OF DIGITAL LIBRARY FOR THE DOD INFORMATION OPERATIONS CENTER FOR EXCELLENCE (IOCFE)

Trisha N. Francis
Lieutenant, United States Navy
B.S., University of Michigan, 1996

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN INFORMATION SYSTEMS TECHNOLOGY

from the

NAVAL POSTGRADUATE SCHOOL
September 2006

Author: Trisha N. Francis

Approved by: Steven Iatrou
Thesis Advisor

LtCol Karl Pfeiffer, USAF
Co-Advisor

Dr. Dan Boger
Chairman, Department of Information Sciences

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

In a memo from Paul Wolfowitz, Deputy Secretary of Defense, “The Naval Postgraduate School (NPS) is hereby designated the DoD Information Operations Center for Excellence. In that capacity, NPS shall facilitate development of Information Operations as a core military competency and innovation.” Commander, US Strategic Command (USSTRATCOM) will serve as Operational Sponsor for the Center on behalf of the Combatant Commands. The Secretary of the Navy and Commander USSTRATCOM will develop a charter for the Center on Wolfowitz’s approval, in coordination with the Under Secretaries of Defense for Policy and Intelligence, the Chairman of the Joint Chiefs of Staff, and other DoD officials as appropriate. The charter will address oversight and activities of the Center, including graduate education, research, research opportunities, and transformation. As a tool to enhance the IOCFE USSTRATCOM is looking into the development of a digital library which will specifically provide resources for the Information Operations Community. This thesis conducts a preliminary requirements analysis for the development of a digital library. Successful development of this digital library is expected to effectively enhance the operational areas of Information Operations and Information Warfare within the Department of Defense

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF FIGURES

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

A. OBJECTIVE

This thesis researches the requirements for the development of an effective digital library for the DoD Information Operations Center for Excellence at the Naval (IOCFE) Postgraduate School. This digital library will be utilized as a research tool and act as a main resource of information on the subject of Information Operations. Electronic resources libraries can develop their collection of Web-accessible information resources based upon the same basic collection policies articulated by libraries or developers for print and electronic resources accessed locally. Most libraries have devoted considerable time to developing collections of materials that best serve their communities of interest. However, some unique criteria exist for digital resource collection.

Selection is a process of comparing individual resources against criteria defined in a digital library's collection development policy, evaluating the quality of documentation, determining the relevancy of the resource to the information needs of your patrons, and deciding whether the library can afford to provide access to a given resource. Design guidelines and evaluation criteria can be employed to build more usable systems but only to the extent that design goals are appropriate for the application. At the core of effective digital library design is the relationship between the content provided and the user community to be served. The content that is contained within the digital library must meet the needs of its users. The design goals can originate from either perspective. Through research of pre-existing digital libraries and discussion with stakeholders the needs of the Information Operations community can be identified and the optimal design guidelines developed for the IOCFE digital library. This community consists of those that are researching, teaching, or learning in any of the core areas of Information Operation. Successful development of this digital library is expected to effectively enhance the operational areas of Information Operations and Information Warfare within the Department of Defense.

B. BACKGROUND

Information Operations (IO) is defined as "The integrated employment of the core capabilities of Electronic Warfare, Computer Network Operations, Psychological Operations, Military Deception and Operations Security, in concert with specified supporting and related capabilities, to influence, disrupt, corrupt or usurp adversarial human and automated decision-making while protecting our own."

IO, by definition, is a compilation of many disciplines. It emerged as a result of the US military's increased dependence on computers and networks. Operations conducted during Operation Desert Storm indicated that technological development had provided the military with computer-based tools and techniques that could be used to degrade not only military systems but those of government and the private sector as well. The development of IO tools and techniques is evolving at a pace equivalent to the rate of technological change within computer industries. The objective of developing the IOCFE digital library is to create a tool which will enable the IO community to keep pace with the rapid changes associated with their ever-expanding area of interest.

C. ORGANIZATION

The main contribution of this thesis is the analysis of user needs balanced against technological capabilities to develop an IO digital library.

Chapter II will discuss the research methodology, specifically the systems analysis and design methods that are applied to information systems and computer applications as they apply to the development of a digital library. These methods will be geared towards the development of a digital library.

Chapter III delves into the analysis of the technology that exists to date. By utilizing digital libraries that exist as templates for the development of the digital library an organization can take the best aspects of each library that are applicable to the development of the IOCFE.

Chapter IV will analyze the rational for building a digital library. This is to include the reasoning for the design, its implementation and the required maintenance that is entail in its upkeep.

Chapter V discusses criteria stipulated by the user community as well as the taxonomy of a digital library. This chapter delves into what requirements need to be met in order for this resource to be a useful tool, an effective digital library that meets the needs of all of its users. In addition, this chapter will discuss that type succinct verbiage that must be utilized in order for the digital library to be understandable and easy to use by all of its customers.

Chapter VI discusses conclusions and recommendations. This chapter briefly addresses some of the additional issues that came up while conducting the study but outside the scope of the thesis. These additional ideas are areas that the author recommends for further study. Some of these issues will need to be addressed prior to to the design/implementation phase of the digital library.

THIS PAGE INTENTIONALLY LEFT BLANK

II. RESEARCH METHODOLOGY

A. DEFINITION PHASE

The definition phase (or project planning phase), identifies scope, feasibilities, schedules, resources and budgets. It is recommended that the scope of the project be agreed upon prior to attempting to identify and schedule tasks or resources to those tasks. Scope defines the boundaries of a project – the parts of the business that are to be studied, analyzed, designed, constructed, implemented, and ultimately improved. It also defines the aspects of a system that are considered outside the project. **[1]** The scope of the project is to research the building of a digital library that would provide resources, offer intellectuals access to, interpret, distribute, preserve the integrity of, and ensure the persistence over time of collections of digital works so that they are readily available for use by a defined community or set of communities. The user community can be defined as those associated with any element of information operations. It is essential that their points of view are known when building this digital library.

The user community entails those individuals that are associated with any element found in the definition of IO. In accordance with the Army Field Manual (FM) 3-13, Information operations is the employment of the core capabilities of electronic warfare, computer network operations, psychological operations, military deception, and operations security, in concert with specified supporting and related capabilities, to affect or defend information and information systems, and to influence decision-making.**[2]**

B. REQUIREMENTS PHASE

The requirements phase answers the question, "What do the users need and want from a new system?" This phase plays an important role in the success of any new information system. New systems should be evaluated, first and foremost, on whether or not they fulfill business objectives and requirements, regardless of how impressive or complex the technological solution might be. [3]

The information needs of the users were gained through interviewing personnel in the IOCFE, those that work within the core areas of IO, as well as the stakeholders

involved in the Homeland Security Digital Library. The main objective of the interviews is to consolidate the wants and needs of the potential users. Those that were interviewed were individuals who are considered subject matter experts in the IO Community. They were contractors, service members of all military services as well as civilians that work in the core area of IO. This gives the prospective designers and developers an idea of what the stakeholders are looking for which can in turn assist them in answering the question, "How the digital library should operate in order to meet the functional needs of its users?"

C. EVALUATION PHASE

The IOCFE wants to research the capabilities and requirements of a digital library that will meet the IO communities' information needs. After analyzing and reassessing the pre-existing libraries requirements, the paper will look into the feasibility of the project. By researching into the development of the HSDL and its purpose, just from initial appearance and depth of detail within the infrastructure of the library some modifications will be necessary in order to ensure that it meets the needs of the customer.

After the determination of the inputs, outputs and processes that will be necessary to meet all of the specifications for a successful library the next step is the design phase.

D. DESIGN PHASE

The design phase, also known as the physical design phase, includes hardware selection, determination of software with regard to custom versus off-the-shelf applications, the design of user interfaces and data capture devices, the specification of data input and report formats, the media to be used for input and output, and the determination and construction of the structure for the corporate databases to be used by the new system. **[4]**

An important part of the design phase will be the user interface which connects the user to the system through menus and forms. Due to the scope of this thesis the different types of design phase implementations and procedures will not be discussed. There are many different types of techniques or strategies for conducting system design, they include modern structured design, information engineering, prototyping, Joint

Application Design (JAD), Rapid Application Development (RAD) and object-oriented design. In this portion of the thesis each type will be addressed along with some of its advantages and disadvantages as it applies to digital libraries.

Modern structured design is defined as a process-oriented technique for breaking up a large program into a hierarchy of modules that result in a computer program that is easier to implement and maintain. Synonyms are top-down program design and structured programming. **[1]** This method is considered a popular technique involving the design of mainframe-based application software and is used to address coupling and cohesion issues at the "system" level.

Prototyping is the process of building a model of a system. In terms of an information system, prototypes are employed to help the system designers build an information system that intuitive and easy to manipulate for end users. Prototyping is an iterative process that is part of the analysis phase of the system development life cycle. **[1]**

Information Engineering (IE) is a model-driven and data-centered, but process-sensitive technique for planning and analyzing, and designing information systems. IE's primary tool is a data model diagram. There are two specific types that will be addressed and considered as possible design tools in this thesis, they are Joint Application Design and Rapid Application Design

- Joint Application Design (JAD), is a technique ensuring that information is gathered from all affected parties, and that requirements that are received in outcome are approved by all participants, and not only by decision of system analysis collecting the requirements. [5] JAD allows users to share their opinions on the current system, and gives a chance through shared purpose to come to a consensus on what needs to be changed. [6] JAD systematizes the systems requirements process, solving project managers dilemma of uniting disciplined approach to systems analysis with flexible user coordination. [7]
- Rapid Application Development (RAD) is a methodology for compressing the analysis, design, build and test plans of a series of

short, iterative development cycles. **[8]** RAD can simply be thought of as a software development process that allows usable systems to be built in as little as 60-90 days, but it tends to come with some compromises.

Lastly, there is the object-oriented design method. Object-oriented design has also been defined as an implementation method which programs are organized in object collections that cooperate among themselves, each object representing an instance of class; each class is part of a class hierarchy and all classes are related through their inheritance relationships. **[3]**

E. IMPLEMENTATION PHASE

This step involves the actual delivery of a product. Due to the scope of this thesis the avenues that will be taken in order to implement the Information Operations Center For Excellence Digital Library will not be addressed in this work.

III. ANALYSIS OF TECHNOLOGY

A. DEFINITION OF A DIGITAL LIBRARY

Despite their increasing popularity, defining *digital library* remains problematic. Of the many definitions that exist, one arising from within the computer and information science research community originated in a research workshop on scaling and interoperability of digital libraries: "A digital library is a system that provides a community of users with coherent access to a large, organized repository of information and knowledge." **[9]**

In contrast, the most succinct definition arising from the community library practice is set forth by the Digital Library Federation (DLF): Digital Libraries are organizations that provide the resources, including the specialized staff, to select, structure, offer intellectual access to, interpret, distribute, preserve the integrity of, and ensure the persistence over time of collections of digital works so that they are readily and economically available for use by a defined community or set of communities." **[10]**

The significance of establishing a precise definition is so that the requirements or needs of two user communities: those who utilize the information (i.e., the IO community) and those who develop and maintain the information (e.g., the Subject Matter Experts (SME), librarians, etc.). As discussed later in this chapter, researchers focus on digital libraries as networked information systems and as content collected on behalf of user communities, while librarians focus more on digital libraries as institutions or services. User communities are those personnel that will be utilizing this resource; hence they are the ones that will play a key part in the overall design. But the librarians must also be taken into consideration; they also play an integral role in the designing and maintaining of the digital library.

When one thinks of libraries, books automatically come to mind. The library has a place in the Western mind as the home, or perhaps even the temple, of the book (and not just because the word library is from the Latin *liber*, meaning "book"). But libraries have held and cared for many other types of materials. Serials (journals, newspapers, and

magazines), audiovisual materials, organizational records, and personal papers all have their place in modern libraries archives, and special collections. And long before the introduction of digital technologies, libraries had to contend with a range of media and technologies well beyond the catalogs, stacks, lamps, and desks that support the use of books. **[11]**

These communities are not mutually exclusive, of course, and most large digital library research projects involve librarians as well as scholars from information and computer science. The aforementioned definitions extend the scope of the digital libraries in several directions, reflecting the contributions of scholars from a dozen disciplines. It tends to move beyond information retrieval to include the full life cycle of creating, searching, and utilizing information. Vice just collecting content on behalf of user communities, it should be a tool that encompasses information-related activities from multiple information institutions. The concept of the defined communities, in the case of the user community will be addressed later on in this chapter.

B. ANALYSIS OF DIGITAL LIBRARIES

There are certain issues that come about when discussing the development and usage of digital libraries. One of the biggest concerns is the advancement of technology and its ease of use. [12] The situation is different with this new community of users. Due to the wide spectrum of users, there is a variance that should be taken under consideration when building a digital library. Users vary by age, by educational background, there is the range of cultural and ethnic perspectives, and then there is the variance in experiences with computer technology. With each different type of user there is a variance in skill level when it comes to technology. Everyone, regardless of there level of knowledge when it comes to computers, should be able to use the IOCFE with ease. One of the criteria that have been stipulated by the user community regardless of the vast difference amongst the users is user friendliness.

The challenges that occur when developing information systems:

- they need to determine what factors make computers difficult to learn and use;
- need to define a set of characteristics for a “user friendly” system;
- need to apply the research to design;

The last item listed above can be applied to the purpose of this thesis; in order to design an appropriate technological system that meets the criteria of its user it is necessary to conduct thorough background research before designing a product. Proper research can mean the difference between having a good product that is usable and having wasted man-hours and money on a bad product.

Users have certain expectations when it comes to systems that they utilize as a tool to make researching easier. Those expectations are **[9]:**

- Easy to learn.
- Easy to use.
- Easy to relearn
- Flexible in adapting to a more diverse user population.

Perspectives on usability have shifted substantially over the course of the 20th century. The initial purposes of ergonomics were to place people into the technological order. Human skills were measured relentlessly so that people could be matched with the machine task to which they were best suited and machines could be operated by those with requisite capabilities. It appeared by the early 1980s, the focus of ergonomics had shifted toward human capabilities and needs. The period seemed to have merged the transition from mainframe computing systems operated by skilled professionals to desktop computing for end users.

Despite all technological advances, establishing generalizable benchmarks for usability remains problematic due to the variety of applications and the diversity of user communities served. Many criteria and guidelines for usability have been derived from the findings of research in human-computer interaction. Perhaps the most general are the

requirements for being fluent with information technology. To achieve the goal of having every given interface that is useful to the population that is utilizing then information technology and especially tools such as digital libraries need to meet specific criteria.

With the research that has been conducted on e-library/digital libraries there was one area that was recommended to concentrate upon, content. [12] Content is the basis and the final result of any research whether beginning an electronic collection development from scratch, or seeking to supplement. The process of selecting quality appropriate materials demands a thorough investigation and attention to detail.

Two unavoidable factors have made it imperative for libraries to build Web-accessible resource collections for their current and potential library patrons – the ubiquity of personal computers, and the publication of information on and through the World Wide Web. **[9]**

The benefits of building a digital library:

- enhances existing services for core user groups;
- provides new services to core user groups;
- attracts new library patrons;
- provides new (or better) services to patrons who are reluctant (or find it difficult) to come into the physical library;
- global information distribution;
- release from physical constraints;

Electronic resources librarians can develop their collection of Web-accessible information resources based upon the same basic collection policies articulated by libraries for print and electronic resources accessed locally. Most libraries have devoted considerable time in developing collections of materials that best serve their communities of patrons.

C. ANALYSIS OF DEVELOPMENT METHODOLOGIES

From all the different types of methodologies that are available the one that appears best suited for the development of the IOCFE digital library is prototyping. The prototype will be the vehicle for developing the full requirements for the system, and its definition should establish the preliminary requirements for the system. Defining the prototype before building it helps users and developers think through the basic functions of the system. The advantages of utilizing the prototyping are **[13]**:

- Reduces development time
- Reduces development costs
- Requires user involvement
- Developers receive quantifiable user feedback with its use
- Facilitates system implementation since users know what to expect
- Results in higher user satisfaction
- Exposes developers to potential future system enhancement

Prototyping is the process of building a model of a system. In terms of an information system, prototypes are employed to help system designers build an information system that is intuitive and easy to manipulate for end users. Prototyping is an iterative process that is part of the analysis phase of the systems development life cycle. **[13]**

What will the prototype initially be and look like? It will be a working model of the system that will include the major program modules, the database, the essential screens, the reports, and the interfacing inputs and outputs used to communicate with other systems. It will be a skeletal version of the system and will not contain all the processing and validation rules that the system will finally have.

Prototyping can come in many different forms – from low tech sketches or paper screens from which users and developers can paste controls and objects, to high tech operational systems using Computer Aided Software Engineering (CASE) or fourth generation languages such as Visual Basic, or somewhere in between. Many

organizations tend to utilize several prototyping tools in order to develop a functional product that meets the needs of the users and is feasible for the developers to create.

The guidelines that are practiced when utilizing prototyping which have been thought to create a successful product are as follows **[13]**:

- Prototyping should be employed only when users are able to actively participate in the project;
- Developers should either have prototyping experience or given training;
- Users involved in the project should also have prototyping experience or be educated on the use and purpose of prototyping;
- Prototypes should become part of the final system only if the developers are given access to prototyping support tools;
- If experimentation and learning are needed before there can be full commitment to a project, prototyping can be successfully used;
- Prototyping is not necessary if the developer is already familiar with the language ultimately used for system design;

The goal of applying analysis and prototyping methodology to the graphical user interface early in the development life is to produce the most reasonable interface within practical business constraints. This occurs by being able to eliminate or revise features and exploit easily supported functionality with limited commitment in terms of time and capital invested. There data provide the analyst with the information necessary to be able to assess which features are critical to the usability or future enhancement path of the application. In the case of the IOCFE digital library development there are pre-existing digital libraries that can be used as template for its development.

The following are some benefits to this early analysis and utilizing prototyping **[13]**:

- Keep ultimate product vision in sight;
- Ability to distinguish between features that are critical and shape the product's future and those features that can be dropped or added incrementally after release;
- By developing rapid and disposable prototypes rather than time consuming code, avoid management feeling committed to use after expending resources;
- With ability to discuss interface behavior with developers implementing each feature, can avoid misinterpretation or oversights by mangers and developers;
- Allows clarification of details missing from functional specification and resolution of design problems before implementation;
- Can develop release criteria that allow decisions to be made regarding added functionality, interface design tradeoffs, and whether product ready to be released;
- Can establish minimum and target goals for specific criteria;

The information element definitions that are utilized for the development of the prototype definition will not be complete and exhaustive. The success of the prototype and the system will depend on whether it is based on a good initial understanding of the information elements. Even though both the users and subject matter experts think they comprehend the information as they begin defining the spectrum of the system, from the interviews and research, which will be presented in the IOCFE Digital Library and User Community chapter, there still resides the confusion in regards to defining the information element.

D. DEFINING WEB SERVICES

The Information presentation on the Internet comes in different modes. The three primary presentation modes are web pages, web services, and web portals. A web page can be compared to a picture. It is a snapshot of information at a given time. This currency of information is dependent on the site administrator to stay up-to-date. The web service extends the idea to provide a service. Retail sales sites are web services. The site administrators control the content of the site, but it is designed around the needs of a user/customer. In order to purchase an item from Amazon.com, the customer must establish an account with relevant information. From there, the customer's information and preferences are recorded in a database for future use. This allows the site to tailor itself to the user, although it is the administrators who decide how the customer's information will be used and what information will be provided. The site may even have dynamic links to other sites and services .

This section will concentrate on web services, the differences between two, three, and four tier applications, and the advantages and disadvantages of each. A web service is a standard approach for making an application available to the outside world. The World Wide Web Consortium (W3C) oversees web services standards and defines web services as follows. A web service is a software system designed to support interoperable machine-to-machine interaction over a network. It has an interface described in a machine processible format, more specifically the Web Services Description Language (WSDL). Other systems interact with the web service in a manner prescribed by its description using Simple Object Access Protocol (SOAP) messages, typically conveyed using HTTP with an XML serialization in conjunction with other web-related standards. **[14]**

Web services are a self-contained, modular application described, published, located and invoked over a network. They are an Internet service that uses the Extensible Markup Language (XML) messaging system, independent of any particular operating system or language. Web services proceed from a human-centric to an application-centric design. **[14]**

Some main components of a Web Service are: **[15]**

- Service Provider – makes the service available.
- Service Requestor – web service consumer.
- Service Broker or Registry – a directory.
- Publish – promoting a service to a registry, allowing it to be discovered and invoked by the requestor
- Find – jointly performed by the requestors and brokers, with the requestor describing what is wanted and the broker delivering the results that best match the request.
- Bind – between the requestor and the provider, allowing the requestor to bind to the service

Web services also support the following attributes: **[16]**

- Reusability – based on the ideal of object oriented design, the code for web services are components that can be reused.
- Loose Coupling – the functionality is isolated from the client and accessible only through an interface.
- Discrete Functionality – is self contained and performs a single task.
- Programmatic Access – intended to be accessed by other programs.
- Internet Accessible – accessed over the Internet using standard transport protocols.

1. Two-Tier

In two-tier (client/server) architecture the computing client talks directly to a server with no intervening process. See figure 1. In a two-tier application, the application program runs on the end user's computer (the client) and communicates with the server (e.g., database server, etc.) through a network or modem connection. In a database

client/server application, the client can pass SQL statements through a Transmission Control Protocol/Internet Protocol (TCP/IP) connection and if necessary, a database specific protocol (e.g., SQLNet for ORACLE, etc.), to the database. The results are returned to the client machine via the same middleware protocols and are displayed to the user. The two-tiered client/server architecture is the most common architecture on microcomputer-based LANs. Hence, the clients manage the user interface, validate the data entered by the user, post requests from clients, execute database retrievals and updates, manage data integrity, and control transactions.

Web tools and databases are technologies that were developed separately, however both technologies are based on two-tiered client/server architecture (figure 1). The partitioning functions between Web browser (client) and a Web server (server) are very instinctive. The Web Server delivers HTML pages and the Web browser displays those pages by interpreting the HTML tags. Neither side can change this division of functions. Because of this simplicity and standardization, many vendors can create web browsers.

When it comes to the partitioning of the functions between database client and the database server, it is much less distinctive. Decisions about partitioning the functions are often made by applications programmers, and are influenced by the requirements of the project. Therefore, there is no standardization.

The typical client-server architecture that is depicted in figure 1 works well in relatively homogeneous environments with fairly static rules. For dispersed, heterogeneous environments with rapidly changing rules, there is client-server architecture, called three-tier client-server architecture. In this type, additional middle tier functionality is added to the configuration.

The main advantage of a client-server database system is that, since the bulk of the database processing is done on the back-end, the speed of the DBMS is not tied to the speed of the clients workstation. Because the client is separated from the server, users are no longer limited to one type of system platform. The clients can be IBM compatible PCs, Macintoshes, UNIX workstations or any combination of these, and run in multiple operating systems.

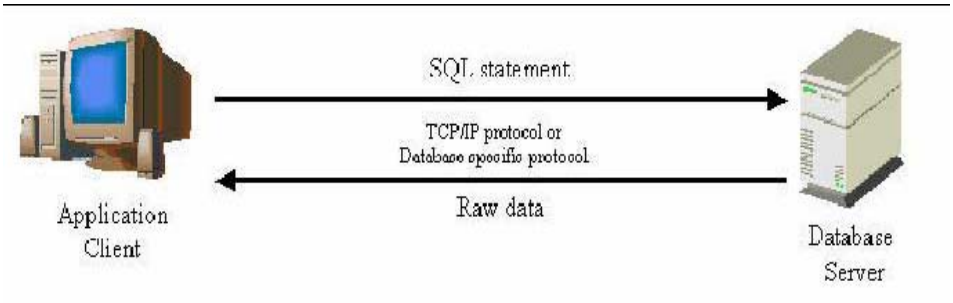

Figure 1. Two-Tier Architecture. [From Ref 15]

However, according to the some research studies conducted by the Standish Group 30% of the client-server projects fail. A common error in client/server development is to prototype an application in a small, two-tier environment, and then scale up by simply adding more users to the server. This approach will usually result in an ineffective system, as the server becomes overwhelmed. To scale to hundreds or thousands of users properly, it is usually necessary to move to three-tier architecture. **[17]**

2. Three-Tier

An important advantage of this architecture over the two-tiered is that it helps clients and servers to process their work. In other words, it allows clients and servers to lose weight and become "thin clients" and "thin servers". This means that the partitioning of functions can be carried further, and greater modularity can be achieved. It is usually agreed that transactions should be implemented in the middle tier. Other processes that could be implemented in that layer are translating data from legacy applications on mainframes, handling security and authentication, and generating reports.

Web database applications combine their two-tiered parent technologies into a new kind of system. This new system is based on the three-tiered client/server architecture (Figure 2). A web browser occupies the client tier, a database server occupies the server tier, and a middle tier holds a Web server and a server extension program. Eventually, this architecture reduces the network traffic, makes components interchangeable, and increases security. However, this architecture also makes database

transaction processing more difficult because of the stateless nature of the HTTP protocol that is used to transfer data between the web browser and the database.

The Web browser sends a web page request or data request to the Web server. The Web server takes the page request and ships the data request to the server extension program, which is connected to the Web server. Then, the server extension program accepts the requests and converts them to a form that the database server can interpret. For the next step, the database server performs a task, such as a query, insert or update, and returns a result set to the server extension program. The server extension program converts the database result to a form that the Web browser can accept (i.e., HTML), and finally it passes the result set to the Web server, which passes the final result to the Web browser.

One of the most important reasons for using a server extension program in the middle tier is to take advantage of the standards that already exist in the two last tiers by translating between the Web server and the database server. Other reasons for utilizing server extensions include handling database connections to reduce the overhead associated with opening and closing the database. Server extensions also support interchangeability at their standard interfaces. Thus Web servers and database servers can be replaced or upgraded with relative ease. This capability could be an essential tool for the development of a digital library.

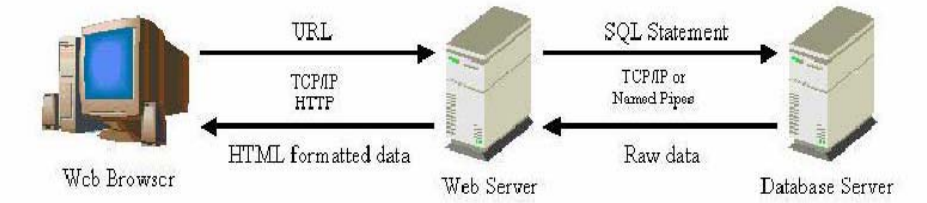

Figure 2. Three-Tier Architecture Diagram. [From Ref 15]

3. Four-Tier

In the four-tier architecture model, the database server remains the data storage and retrieval mechanism and the application server continues to act as the container for implemented business logic. The presentation server becomes the contact point for the client. All requests and responses originate there. The client can insert a structured XML payload, consisting of commands, data, etc. into the presentation server request allowing for structured payloads rather than a flat unstructured payload in name value pairs. The response sent by the presentation server to the client contains metadata for the audio and video content. The actual audio and video is still served by systems specifically made to serve those content types, taking advantage of their particular performance tuning. The application server has access to the audio and video content servers to facilitate its interaction with the meta-data of those media types. The presentation server can communicate with the application server using any combination of naming services and remote interfaces (such as the RMI/JNDI/IIOP combination.) This can also be done through a stateless protocol, using HTTP alone, or with a high layer protocol such as SOAP, XML-RPC, etc.

The presentation server will perform one of two actions: a) retrieve cached content and return it to the server or b) send one or more requests to the application server and/or other content servers. If cached content is available and appropriate, it is returned. If cached content is not available, one or more requests are made to the application server, which performs the necessary business logic and returns a response to the presentation server. The presentation server formats the data received according to the presentation logic, the client device capabilities, its user settings, and content type. This is sent to the client, which retrieves the described content. The presentation server abstracts all presentation logic from the client and the application server, and sends the appropriate data when it is requested. It performs all presentation formatting before the response is sent to the client. **[12]** See Figure 3 for a diagram of this configuration. Right margin should be justified – check all the way through -

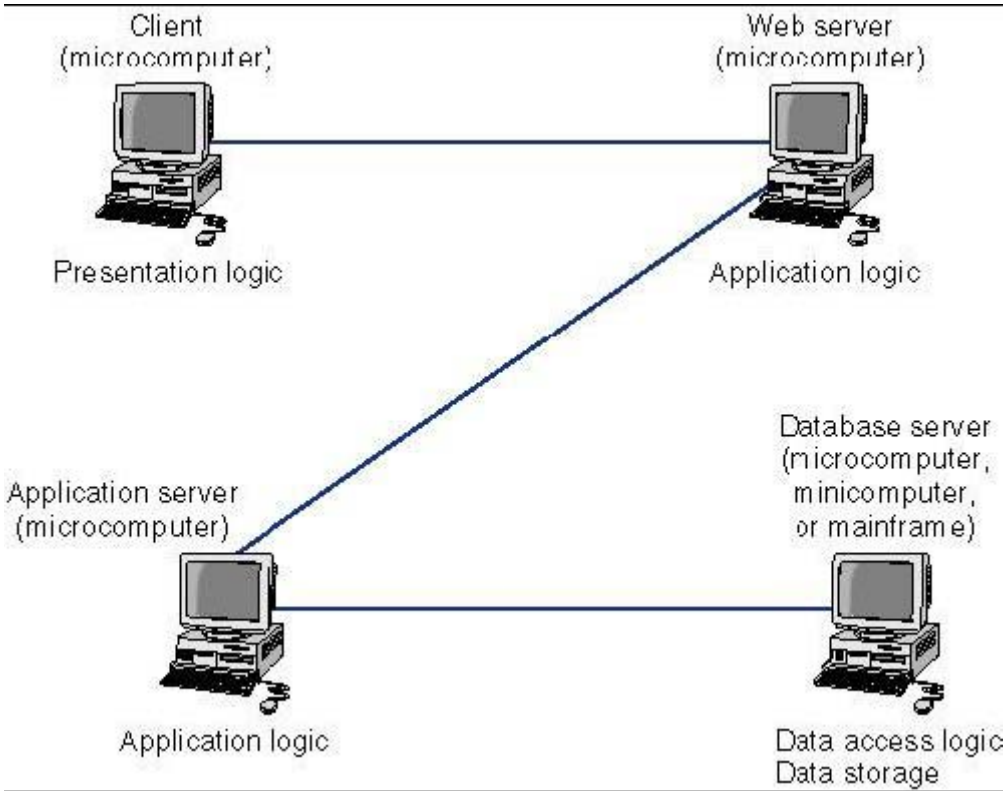

Figure 3. Four -Tier Architecture Diagram. [From Ref 18]

IV. PRINCIPLES FOR BUILDING DIGITAL LIBRARIES

Building a digital library is expensive and resource-intensive. Before embarking on such a venture, it is important to consider some basic principles underlying the design, implementation, and maintenance of any digital library. Digital libraries include personal, distributed, and centralized collections such as on-line public access catalogs and bibliographic databases; distributed document databases; scholarly and professional discussion lists; and electronic journals, other on-line databases, forums, and bulletin boards. In the case of the IOCFE digital library, the digital library must entail a digitized collection which provides a full text of materials, it must be accessible through the Internet, and it should be indexed, searchable or browse-able in a consistent manner. **[11]**

The purpose of a digital library is very similar than that of a physical library. A physical library has been described as four-fold; collection; organization and representation; access and retrieval; and analysis, synthesis, and information resource. **[19]** The resource is used by a client-population and for cost-effective storage and preservation of such resources. Organization and representation have to do with classifying and indexing information resources in ways relevant to their potential users. Access considerations include design of physical space and organization of materials within such space to respond effectively to user needs and expectations. Information retrieval has been addressed, of course, in the design of systems specific to that task. Analysis, synthesis, and dissemination functions include responding to reference questions, and producing evaluative reviews.

Despite the diversity and often conflicting goals and objectives, the Internet community seems to have collectively grasped that mere connectivity to electronic resources cannot guarantee utility or satisfaction. Without effective information management, the Information Superhighway will remain an unpaved dream. The search for an appropriate model, upon which to base the enormous task of restructuring the world's stockpiles of data resources, requires the uncovering of the overlooked, and often unappreciated, field of Library Science. It could be possible that just by looking into the classic Library Science that a good working model for the web can be designed.

Though a promising candidate, at issue was the adaptability of library technologies and practices from the realm of maintaining on-site collections of physical media to the management of remotely stored, electronic resources. While preliminary results from several Digital Library research projects confirmed that the principles of Library Science could be applied to the world of electronic media, they identified a significant void in the capabilities of existing information-related technologies. In 1994, several countries, including the United States, committed their resources to numerous, large-scale, well-funded Digital Library Initiatives. Within a few months, these programs were joined by hundreds, then thousands of local development projects aimed at bringing yesterday's academic, public and private libraries into the 21st century. Each of these programs has self-motivated goals, but together they contribute to a world-wide Digital Library. Movement that is collectively expanding the horizon of technology and science. **[20]**

A. FUNDAMENTAL ISSUES

The topic, Communications, highlights some of the many projects underway for the creation or enhancement of digital libraries. At the moment, no one seems to think there will be only one gargantuan digital library to satiate the public's appetite for information. Rather, the expectation is that there will be many digital libraries, most of which will have specialized collections and will be networked together in a way loosely resembling today's Internet. **[21**] Electronic access to an almost unfathomable quantity of data has been facilitated by huge strides in both the technology and availability, at low cost, of communications connectivity. This trend should continue, though not without difficulty. A major obstruction to the attainment of on-line accessibility to remotely stored data is the requirement for both the user and provider to establish compatibility through standardization, standardization not only through applications but presentation as well. Yet, connectivity and compatibility are only two of many challenges that must be overcome before information can efficiently be shared around the globe.

1. Data vs. Information Resources

A short trip on the Information Superhighway via an Internet web browser demonstrates the point. One of many powerful Internet search engines can use a key word or phrase to sift through thousands of remote sources and deliver to the user a list, of potential candidate items. The information-seeker is confronted by a data collection whose size, completeness, accuracy and utility are determined by chance. Information technology and digital libraries help alleviate this problem. In a test conducted at NPS on 15 Oct 1995, a search conducted using the key word "Pentium," resulted in a list of 947 sources whose composition spanned the spectrum from technical material, to media reports, to humorous articles and personal opinion. While sifting through this pile, there were hundreds of duplicates, dead-ends or nonsensical sites that took many hours to eliminate. A lesson learned from using the Internet is that it is relatively simple to accumulate mounds of data, but chasing down valuable information is a non-trivial task. **[20]**

Clearly, connectivity is a double-edged sword that, while useful in rounding up potential sources, can cut deeply into one's time budget and still provide a less than satisfactory result. This is encountered on the Internet daily, by millions of information-seekers, and is magnified by users that are not physical located near the library source, those who cannot afford to waste precious time, or have limited bandwidth, that are in pursuit of solutions to crucial problems. It is the demand for efficient navigation, selection and retrieval of information, from millions of remote data sources, that has sparked the Digital Library movement **[22, 23, 24].**

2. Data Structuring

Information is data transformed by format, filtering, analysis and/or accessibility into a product that has value to the user. To facilitate this capability, a would-be information provider must accurately forecast user needs, employ a robust organizational method and be committed to diligent maintenance. One approach, frequently used for large databases, involves the creation of metadata, which is a separate data-set that provides complementary information on the structure, organization, and content of resources, but does not require the cache of the resource itself**. [25]** Similar to a library

card catalog, metadata contain a relevant description of the source and material while providing the information-seeker with a convenient environment to search.

Given quality metadata, there still must be an effective process to interface both user and provider (with adequate security), and functionally isolate and extract the desired information from the data store. Then there must be a suitable mechanism to transfer the product without compromising its integrity. With such a system, a pool of trained users could conceivably tap, search and exploit this one data resource. The reader should gain some appreciation for the magnitude of the challenges facing the digital library movement by imagining this effort compounded by millions of potential digital library users and data resources, eventually integrated into a "user-friendly," world-wide system. **[20]**

B. INFORMATION ACQUISITION

The level of effort required to electronically search, locate and capture valuable information is not simply a function of data rate. It is determined by the structure of the data collection, the quality of its indexing, the power of the search and retrieval system and the expertise of the user.

Internet searching is metaphorically like casting fishing net. Without knowledge of the form, density and distribution of the objective, the composition and quality of the "catch," is strictly up to chance. In the world of digital data, this means that the info-seeker must manually sort random results, which can range in utility from useful to absurd. The cost in time alone can be enormous and there is no guarantee that an exhaustive search has been accomplished. To solve this problem, the DL community is debating a new electronic information management paradigm which contrasts *two* dissimilar approaches to capturing information: The Library Approach**,** which replicates the environment and the related processes of a physical library; and the Unstructured Approach**,** which embodies the information search and retrieval techniques used in wide practice on the Internet today**. [26, 27, 28]**

1. Library Approach

Librarians have established a system that consistently satisfies the differing information needs of a widely disparate user group. This has been accomplished by structuring physical media (data) into logically organized and accessible collections and providing extensive cross-referencing through cataloguing and indexing (metadata). A library supports an information search strategy focused upon:

- Evaluating all valid, available sources for candidate items;
- Quickly and automatically eliminating alternatives;
- Acquiring for review only the minimum number of items required to accomplish the task; and
- Providing a feedback channel from user to provider.

Some librarians contend that failure to follow such a strategy results in time delays, incomplete research, storage problems, and increased costs. These are precisely the reasons that led the Digital Library community to apply Library Science to the realm of electronic data resources. **[29]** In the environment of physical media, librarians have become so effective at their craft, that library customers universally expect to have their information needs met swiftly, effectively and with minimum fuss**.** Peter Graham, the Electronic Resources Librarian at Rutgers University, in his article "Requirements for the Digital Library," discusses the necessity for applying the structured approach of library science to the inter-networking environment:

> Users' needs will continue to be what they long have been. Users will want information reliably locatable, so that when they go there (whether personally or on the net) they can expect to find what they're looking for. Users will want information easily accessible: the cataloging must be clear and accurate, and the information must be promptly retrievable. Users will expect information to be available that was placed in the library's care a long time ago; and they will expect that the integrity of the information they get from the library will be assured. [22]

Unlike a library, where information is targeted with great precision, Internet accessibility to electronically stored information currently follows a different strategy, the unstructured approach.

2. The Unstructured Approach

Contrast the organized and supportive environment of a library with the lack of structure one encounters on the Internet today. Though early Internet users heralded its freedom from restriction and regulation, the Internet's explosive growth has brought it to the brink of information chaos. There is such a huge amount of information yet no uniformity of how it is collected or categorized in order to allow for easy means of search.

When searching for information, most users set an arbitrary limit on the number of items displayed on-screen, which indiscriminately filters most of the candidate sources because of time constraints. It is doubtful that many individuals routinely inspect sites that have been listed beyond the display limit. What remains is a hodge-podge of topics, linked only superficially by the existence of a key word or phrase. The user is left to wade through this jumbled mess as thoroughly as his or her time and patience will allow.

If a likely candidate for electronic transfer (download) is found, the possibility of successful capture and future utility is dependent upon format comparability and user expertise. In most cases there is no guarantee of accuracy or authenticity for the user. Compounding the confusion are millions of user-generated linked-lists which provide pointers to someone's "favorite" sources. In this situation, the reference is likely offered by a well-intentioned, but untrained person who may be providing misleading or erroneous information. Moreover, these personal lists are erratically maintained and rapidly become outdated. Without standards for cataloging and indexing, and given the disparity between user expertise and interests, the Internet landscape has become a maze of conflicting signposts and is replete with duplication, nonsense links and inactive sites. **[20]**

For users who face connectivity charges, the problem is magnified. Evaluating candidate items on-line is expensive and time-consuming. Other problems include:

- There is absolutely no assurance that an exhaustive research on the topic has been accomplished by the user.

- The quality and accuracy of available material varies from excellent to ridiculous.
- Specificity in search criteria is limited by the lack of standards and technology to index and catalog distributed digital material.

To combat these problems, computer and information system specialists and librarians are teaming up to develop full service Digital Libraries which "...accomplish all essential services of traditional libraries and also exploit the well-known advantages of digital storage, searching, and communication**." [29]**

THIS PAGE INTENTIONALLY LEFT BLANK

V. IOCFE DIGITAL LIBRARY AND USER COMMUNITY

A. USER COMMUNITY

A science library in a research university serves multiple clienteles, with varying needs and expectations. Research faculty and graduate students need access to high quality research collection and materials relevant to their area of interest. Timely access to key journals in their fields of research is essential. In this era of great change in libraries, daily collection development tasks and decisions require thoughtful consideration of theoretical issues. Understanding evolving new technologies and considering their impact on building digital collections is necessary for effective collection development. Two strong forces are driving the day-to-day development of a digital library: rapidly advancing technological capabilities and ever-increasing expectations of the users. **[30**] This concept is relevant to the IOCFE digital library due to the need for a research tool that is for a user community who needs access to the broad range of subjects associated with information operations.

Digital libraries design is particularly challenging because human information behavior is complex and technologies are rapidly evolving. Two important aspects of human-centered design are assessing human information needs and the tasks that arise from these needs and evaluating how the digital library affects subsequent human information behaviors. The human-centered design principle links three clusters of constructs or facets – (1) people and their needs, characteristics and contexts; (2) design, implementation and evaluation; and (3) digital libraries. **[29]** Given the complexity of human information needs and the uncertainty about the effects of new systems, multiple data viewers are necessary to guide design and to help understand the impact of digital libraries.

1. Knowing the User

The degree that a digital library design will be tailored to a particular user community will depend on the goals of the application, the profile of the community, the amount of user participation in design and the characteristics of the application. If the scope of the user community is well-defined (i.e., employees of a company or students in

a university or research setting) then a representative sample can be studied, and design participation can be solicited or appointed by designers of the library. Research on primary and perhaps secondary target audiences may identify common elements and requirements as well as the degree to which their needs and interests vary. Such studies can provide a starting point for the design process. One of the largest complaints that have come out is that the designers/developers do not obtain the users input. The frustrations of a lot of users is the front-end is extremely simple yet the back-end of the system is too technical. **[31]**

In addition the issue of content is very important to the potential users. The quality of the content can only be as good or bad as the person or office charged with maintaining the material. If this is an additional duty for the librarian then the quality will suffer. Unless someone is charged with the full-time librarian duties for the database, then an endeavor of this magnitude will go unused. An IO digital library requires research, maintenance, upgrades, question and answer, and adequate network access for each security classification. In regards to security classification, the concept of "need to know" comes to light. The scope of this thesis does not deal with the security classification specifically, yet the issue of how to maintain "classified" and "unclassified" documents will come about. Maintenance of "classified" vice "unclassified" documents entails more detail, depending upon which type is under discussion. With this detail requires management by specific personnel with specific expertise, which entails more costs that are accrued. It costs more to maintain classified documents than unclassified due to the fact that "classified" documents must be kept in special secure spaces, and have special rules for maintenance, distribution and collection.

According to interviewee Scott Runyan, a contractor, the content organization is a volatile subject.

> Many purists will say the information should be organized by the Joint Doctrines – end of subject. Others will state that Information should be divided by each services requirement or cataloged by subject. My preference is all of the above. With search engines and data-mining techniques all are possible. The documents should be mined for keywords with a user-friendly interface. It has been my experience as an online course designer, that information availability (on-line) takes a life of its own – kind of like free market economics. **[32]**

2. Individual Differences

Another consideration in designing digital libraries is the range of skills, abilities, cognitive styles and personality characteristics that are found within a given user community that may affect usability. Collectively these factors are known as individual differences. Population characteristics known to influence usability of digital libraries include computer skills, domain knowledge and familiarity with the system. Other influences include technical aptitudes such as reading and spatial abilities, age and personality characteristics such as those measured by the Myers-Brigg test. **[19, 33]** Users needs are based upon their roles. Despite the published definition of IO in accordance to the Army Field Manual (FM) 3-13, there are some users within the IO community that view IO in a different manner and therefore expect different things from an IO library. According to one of the interviewees, the IO discipline consists of:

- Joint/Combined Air Operations Center or front-line;
- Intelligence and Analysis
- IO Sub-Disciplines
    - Electronic Warfare (EW)
    - Influence Operations
    - Network Warfare
    - Integrated Control Enablers (ICE)
- Joint or Command Specific
- Service Specific **[31]**

From the interviews that were conducted via teleconference, when differentiating between the necessities and the wants of the user community it is very difficult. There are no specific criteria that can be narrowed down, as stated previously users wants are stipulated by their jobs needs. They varied from wanting to see a IOCFE Digital Library as a Element of Power, a resource that can be utilized by all, those involved in research, learning/training of the IO community as well as allows linkage to US allies' doctrines, in order to understand the spectrum of thinking in regards to Information Operations; To the

digital library having to be an “interagency” resource. Some see it as a resource to have on the multi-national level. Users have expressed the need to have doctrines accessible that have been written at the Combatant Command level, as well as have a means to have access to academic research that is on-going or being implemented at that moment in time.

The range of users that are proposed to have access to this library is vast. More knowledgeable users understand that and have insisted that the library be flexible, the availability of information be up-to-date and easily accessible. One of the major complaints that users have found is the ability to find the information that they want. Developers from time to time tend to forget that the hardest part about gaining the necessary information is the mere way of searching. According to Dennis Murphy, AWC, Center for Strategic Leadership, if it takes a user more than 15 minutes to find the information that they are researching, then this resource is not seen as feasible to use. **[4]** It is pertinent that when developing this digital library it is kept in mind that the interface makes it easy to maneuver around the site in order to find information. One interviewee suggests that a helpers guide for the users be available, so that the complexity with the site is made simple. Another suggestion was to offer a tutorial for the first time user. It is important that developers and designers ensure that the digital library is geared towards all of its users. Yet, it tends to be difficult to assess what functional requirements stipulated by the vast user community is more important, therefore the attempt to meet all requirements can lead to staggering costs. Runaway costs are endemic in major projects such as digital libraries. In order to limit huge costs designers weigh the benefits certain functions yielded by the system can provide, those that appear more beneficial are implemented into the system. Costs, especially with information systems technology, at times can be underestimated while benefits are overestimated. **[34]**

Some of the general ideas that were expressed by the potential users is that the library should have a means of addressing and posting lessons learned, means of obtaining experience from one Area of Responsibility (AOR) to another AOR, as well as means of cross communication, this specifically pertains to allowing those that are out-to sea having accessibility. In addition, the users that it was pertinent that once the library is built that there is a subject matter expert readily available when it comes to the content of

the library. This way when someone is unable to find the information that they are seeking on their own, there is someone that they can speak with to assist them in their search. The users that were interviewed have stated that there should be a significant amount of support for a resource of this nature due to the fact that IO is an emergent field, ever-changing, and those that are involved in this evolution need to have a means in assisting in its advancement.

B. LINGUISTIC DISCONTINUITY

Much of the information in the digital library will consist of documents and representations in natural and controlled language. Problems with this include not only the intrinsic problems posed by the ambiguity of language used in a given database, but also by both the quantity and heterogeneity of the information that will be searched and integrated across multiple collections. The identification and construction of these linguistic techniques builds upon the prior research in manipulating surface structure of documents and queries to build synthesized linguistic capabilities into an information-retrieval system. **[35, 36, 37]** These methods make use of existing surface structure found in documents and queries, as well as the structure and content available in already existing controlled vocabularies.

Taxonomy is a principle of classifying living organisms in specially named categories based on shared characteristics and natural relationships. **[38]** Application of taxonomy facets after a search allows users to drill down to meaningful data from a result set. Building taxonomy can be a daunting evolutionary process. According to one of the key personnel, who is specifically involved in the development of the Homeland Security Digital Library's (HSDL) taxonomy. According to a presentation that was given by key individuals in the development of this taxonomy, its evolution went through several stages until there was one that seemed to fit into the scheme of things. "What for us, began as a flat portal hierarchy, was expanded to a thesaurus, and then grew into dimensions that became multi-faceted taxonomy. The facets are incorporated in an ontology to be used for auto-categorization." **[34]** Central to federating any collection of independently-generated information sources or databases is a common language for describing content without detailed information about access mechanisms, organizations,

or any other implementation-specific issues. According to the members interviewed who are part of the HSDL taxonomy team, they have learned that each stage is fraught with its own set of costs and controversy. **[34]**

After developing their base taxonomy manually with the help of a thesaurus editing program, they acquired categorization and search engine tools which were integrated to index the full-text and external metadata of over 20,000 key documents that were manually selected, as well as other content that were harvested through automatic means. According to Pitts and Woon, from the start of the project, they could not decide when to assign a subject or keywords to content due to the fact that their taxonomy was underdeveloped. They realized that maintaining consistency in subject analysis in the new and interdisciplinary domain of the homeland security would be very difficult to achieve. **[34]**

Their goal was to apply automated tools that would facilitate consistency in assigning taxonomy terms to content, yet flexible enough so that changes in terminology and concepts would be easily implemented. Initially, the Knox Library was trying to achieve the grand vision of National Strategy, but instead their focus revolved around the need of their intended audience. When they began their project in 2003, they started by structuring a taxonomy that would enable the Center for Homeland Security's graduate students – local, state and federal civil service policy makers – to access key academic, policy, and news material on homeland security topics. But then in 2004 they expanded their user population to policy makers/practitioners throughout the Department of Homeland Security. Ultimately their goal of taxonomy became to enhance portal browse and search functionality, as well as to aid in their overall content evaluation by identifying gaps in their content.

The development of query paradigms that allow users to retrieve the desired material with ease by processing complex requests in this distributed environment is a key research problem. Traditionally, query optimization techniques determine a fixed execution strategy for a query by evaluating and comparing all information given in metadata. For example, in the University of Michigan Digital Library (UMDL) this will become a much harder problem because the query optimizer will have to make decisions with incomplete information. According to the research that has been conducted on the

UMDL, the primary characteristics of a digital library are that it should provide physical and intellectual access to a highly distributed, heterogeneous collection of information resources. Access should be independent of time and distance, and should be flexible and personalized to the individual. Ultimately, it should facilitate new, collaborative ways of learning, gathering information, and doing research. The University of Michigan Digital Library Project is investigating methods of achieving these goals through a distributed, federated architecture, utilizing designers that embody knowledge about collections, users and query processing methods, as well as mediation procedures to coordinate interactions among them. Their goal, similar to that of most digital libraries' are educational and research enrichment is to effectively guide the user's search toward the best available resources, and avoid the problem of overwhelming the user with too much information. **[26]** One of the integral portions that lead to this accomplishment is to have a well established taxonomy.

While building and implementing taxonomies to facilitate access to information can be a time consuming and expensive task, in the long run, an investment into such tools to work out semantic rules, to enable accurate categorization, will assist in the scaling up of the digital library's ability to provide customized access to ever increasing volumes of content and an increasing diversity set of users.

THIS PAGE INTENTIONALLY LEFT BLANK

VI. RECOMMENDATIONS AND CONCLUSIONS

A. SYNOPSIS OF WORK

This thesis explored the requirements for establishing an effective Digital Libraries. One of the major concepts that needs to be addressed when developing a digital library is taking under consideration that delineating requirements means the difference between success and failure. When relying upon the user community, as far as what specifications the digital library must have as well as the expectations of the digital library, it is precise to state that this process in of itself is a large and tedious one. Simply due to the fact that users really do not understand the concept of digital library, or its meaning. A digital library often gets confused with the ideal that it is simply a repository of documents or that it is a website which in of itself contains multiple links to different sources of information; A digital library is much more than that. As information seekers, the ever increasing access to electronic resources has defined a necessity for new Information Management practices and technologies. In response, the principles of traditional library science are being adapted from the local control of physical media to management of distributed electronic resources. Globally, thousands of ongoing digital library initiatives have been undertaken since 1994, governments, academic institutions and corporations are all contributing to this emerging field. Generally a physical library has an easily identifiable clientele (e.g.. specific university faculty, a specific student body, etc.) A library on the internet inherently has a global audience. By compiling the users needs, an online digital library has the potential to become the tool of the future. It will ensure that when a user with a specific interest has a resource that can be utilized that is easily accessible and user-friendly.

As was shown in the development of the HSDL there is a considerable risk in misidentifying user requirements. This can lead to expensive false-starts and the construction of a system incompatible with its user community. There is also the issue of consolidating all the ideas of the user community, which is vast in experience and knowledge, thereby enabling the designers to build a system that satisfactorily meets the needs of all its users. Although the system may not meet the specific needs of all users it

can, through proper requirements analysis, adequately satisfy its customer base. The object is to ensure that the basic needs of the users are met, some "wants" may be left out yet the sole purpose is to meet the functional requirements, as many as economically and technologically feasible.

B. SUMMATION OF FINDINGS

This thesis merely scratches the surface of the intricate detail and requirements that goes into the development of digital libraries. One of the biggest issues seen across the board when it comes to building digital libraries is the delineation of requirements stipulated by the user community. This is due to the fact that there is such a wide variance in the type of user and their computer capabilities. Once a consensus has been established, next there is the obstacle of the design. Designing the system to meet the criteria, yet ensuring compatibility with other systems, is yet another obstacle. By utilizing pre-existing digital libraries as templates the designers/developers will have examples to give them ideals for their product, stemming from the GUIs to the intricate of the infrastructure.

To date, US military services have concentrated their efforts on the management and control of tactical information. By its nature, this field of work, IO, is extremely security conscious, which in turn, encourages isolation and inhibits flexibility. It is to the contention of the author that a large portion of the military's daily information needs, as service members, are non-tactical in nature and unlikely to be well supported by the tightly controlled combat information infrastructure.

There appears to be a movement towards the increased usage of digital libraries. Through the use of digital libraries in the military, it could come to represent a unique opportunity to meet non-tactical military needs. A lot of times service members cannot gain access to pertinent information due to the fact that either they do not know where to begin their information searches, or it may not exist in digital form in the world wide web. Service members need readily accessible information as well as a method which will enable them to process information quickly in order to conquer tomorrow's challenges. By committing to the development of its own digital library, the Information

Operations community can establish a conduit through which they can influence policy, exploit technologies and tap a limitless amount of resources.

A key role in the development of the digital library is the link between what the end-users want and need, while at the same time providing enough information to the developer so the correct system is developed. Another issue that most of the end-users agreed upon is that the system be user friendly, all that use the system no matter what their technological background may be should be able to use this system; whether with some sort of tutoring or not.

C. RECOMMENDATIONS

The following areas require further study and attention:

1. Make a decision based upon the cost-benefit analysis whether this endeavor is worthwhile to explore.
2. Identify potential funding sponsors/advocates and establish liaison.
3. Identify potential members of the IOCFE Digital Library.
4. Define configuration for initial system platform.
5. Design and establish an IOCFE website based upon user community stipulated criterion.
6. Cross examine other digital libraries such as HSDL, University of Michigan Digital Library (UMDL), or University of Illinois Interspace Project as potential models for some aspects of the IOCFE Digital Library.
7. Seek guidance and support pertaining to DoD specific digital library issues from Defense Information System Agency (DISA) and Defense Technical Information Center (DTIC). This is key in reference to making decisions upon who should have access to the digital library, and what specifications this access should be based upon. In addition, it is integral that the library is accessible to those out-to-sea.

8. Seek potential corporate partners for related research. It is important that not only all services should be able to use the digital library and contribute to its expansion.

D. CONCLUSIONS

Though the concept of digital library is not new the creation of such a tool is not something to be taken lightly. Presently, there is no tool of this magnitude that addresses specifically the core areas of IO. This study laid the groundwork for an effective requirements analysis as a starting point to designing and developing a useful information tool for the Department of Defense. The requirements analysis phase enabled the users and designers to get together and compile a list of agreed upon functional needs and desires that need to be met by the digital library. If a requirement's analysis had been conducted prior to the creation of HSDL, much of the refurbishment that is now underway could have been avoided. According to some of the HSDL personnel there are still yet some changes that are being made, due to the new demands of the HSDL.

Throughout history, the means in which a library maintained its vast amount of information has transitioned. It was merely 20 years ago that people were still using a card catalog, in which things were categorized by name of author, title, and subject. As time has progressed the listings moved from the card catalog to the computer. In the meantime, along came the Internet and with it a new way to find information without having to leave the comfort of your own home. Now the key is to have the ability to have a centralized location for specific information vice having to search the entire web for days at a time, trying to find that "right" site with the latest and greatest information.

LIST OF REFERENCES

1. Whitten, Bentley, & Dittman. Systems Analysis and Design Methods. 6th ed. (2004) McGraw-Hill Companies, Inc. New York, NY.
2. Headquarters, Department of the Army. Information Operations: Doctrine, Tactics, Techniques and Procedures. Nov 2003.
3. Wood, Michael R. "A New Twist". CIO Magazine, 1 Sept 1995. http://www.cio.com/archive/090195/ins_content.html.
4. Marakas, George M. (2006) Systems Analysis Design: An Active Approach (2nd ed) University of Kansas: McGraw-Hill Irwin.
5. Christel, M., Kang, K. "Issues in Requirements Elicitation". Carnegie Mellon Software Engineering Institute. http://www.sei.cmu.edu/publications/documents/92.reports/92.tr.012.html, 11 May 2005.
6. Ulrich, William M. "Getting System Specs Right For The 'E' Era". Computerworld, 19 Mar 2001. http://www.computerworld.com/storyba/0,4125,NAV47_STO58692,00.html
7. Keen, Peter G. "Let' Put Project Management Out Of Its Misery" Computerworld, 7 Dec 1998. http://www.computerworld.com/cwi/story/0,1199,NAV47_STO34677,00.html
8. Creative data – building worlds of information. http://www.credata.com/research/methodology.html, 16 Apr 2005.
9. Borgman, Christine L. *Digital Library Use: Social Practice in Design and Evaluation.* Edited by Bishop, Van House, and Buttenfield. Cambridge, Massachusetts. The MIT Press. (2003)
10. Jeng, Judy, "What is the Usability in the Context of the Digital Library and How Can It Be Measured?', *Information Technology and Libraries*; Jun 2005; 24,2;ABI/INFORM Global pg 47.
11. Iatrou, Steve, interviewed by LT Trisha Francis, Monterey, CA, Dec 2004.

12. Scott Runyan, Contractor 39IOS/DOT, interviewed by LT Trisha Francis, Monterey, CA, 11 Nov 2005.

13. http://www.umsl.edu/~sauter/analysis/prototyping/proto.html, 16 Apr 2005.

14. Cerami, Ethan, Web Services, 2002.

15. SYBEX. SYBEX, ASP, ADO and XML Complete, SYBEX Inc. 2001.

16. Bieter, Kurt, Kosuke, Ishii, Integrating Productability and Product Performance Tools Within a Web-Service Environment, Stanford University, (http://mml.stanford.edu/Research/papers2003/ASME/CIE2003.48281.pdf(Jul04), 08 Feb 2005.

17. Digital Library: Gross Structure and Requirements (Report from a Workshop) IBM Research Report RJ 9840, May 1994. Submitted to IEEE Computer Society Press, Proc. Workshop on On-line Access to Digital Libraries.

18. INFRCOM, Distributed Data Processing PowerPoint Presentation (http://Cui.Unige.Ch/Osg/Courses/Infrcom/lectures/29(Sep04), 08 Feb 2005.

19. Borgman, C.L., D.O. Case, and C.T. Meadow.1989. The Design and Evaluation of a Front End User Interface for Energy Researchers. *Journal of the American Society for Information Science*, 40(2), 86-88.

20. Norris, Robert E. & West, David. The Digital Library Phenomenon: Opportunities and Implications for the Naval Service. Naval Postgraduate School. Monterey, CA. (Thesis).

21. Samuelson, Pamela. Copyright and Digital Libraries. Communications of ACM. Apr 1995. Vol 38, No. 3.

22. Graham, Peter S., *Requirements for a Digital Research Library*, Jul 1995, (http://aultnis.rutgers.edu/texts/DRC.htm).

23. Lesk, Michael, "The Digital Library: What is it? Why should it be here?" *In Source Book On Digital Libraries*, E.A. Fox, ed., Blacksburg, VA: Department of Computer Science, Virginia Tech University, TR 93-95.

24. University of Missouri-Louis, College of Business.

25. Marshall, Catherine, P.J. Nurnberg, R. Furuta, J. Leggett, and F. Shipman III, "Digital Libraries: Issues and Architectures," *Proceeding of Digital Libraries '95 Conference*, (Austin, TX, Jun 11-13), (http://www.csdl.tamu.edu/DL95/), 24 Feb 2005.

26. Birmingham, William P., K.M. Drabenstott, C.O. Frost, A.J. Warner, K. Willis, *And The University of Michigan Digital Library: This is not your Father's Library*, 1994.

27. Fox, E.A., R.M. Akscyn, R.K. Furuta, and J.J. Leggett, "Digital Libraries." *Communications of the ACM, Apr 1995, Vol. 8, No.4.*

28. Levy, D.M., and C.C. Marshall, "Going Digital: A look at assumptions underlying digital libraries," *Communications of the ACM, Apr 1995, Vol 8, No.4*.

29. Bishop, Ann Peterson, Van House, Nancy A., and Buttenfield, Barbara P., "Digital Library Use: Social Practice in Design and Evaluation", MIT Press, Cambridge, Massachusetts, (2003).

30. Gladney, Henry M., and E.A. Fox, *Digital Library: Gross Structure and Requirements:* from a Mar 1994 Workshop.

31. Barnes, Susan, "Becoming a Digital Library", Marcel Dekker, Inc. (2004).

32. Dennis Murphy,last name first AWC, Center for Strategic Leadership, interviewed by LT Trisha Francis, Monterey, CA, 6 Nov 2005.

33. Demiryurek, Ugur, Turkish Navy "Analysis, Design and Implementation Of A Web Database With Oracle 8I" (Thesis) Naval Postgraduate School, Mar 2001.

34. Egan, D.E.1988. Individual Differences in Human-Computer Interaction. In M. Helander, ed., *Handbook of Human-Computer Interaction* (pp 543-568). Masterdam: Elsevier.

35. King, John Leslie and Edward L. Schrems. Cost-Benefit Analysis in Information Systems Development and Operation. Computing Surveys, Vol 10, No. 1, Mar 1978.

36. Warner, A.J. 1992. A Linguistic Analysis of Variant Expressions in Medline. Proposal funded by the National Library of Medicine.

37. Warner, A.J., and Wenzel, P.H.1991. A Linguistic Analysis and Categorization of Nominal Expressions. In *Proceedings of the 54th ASIS Annual Meeting*, (Washington, Oct) J-M. Griffiths, ed. Medford, NJ: Learned Information, 186-95.

38. Pitts, Angela and Woon, Lillian. Session C303: Taxonomies and Classifications: Oh My! Integrating Taxonomies with Search and Categorization. (Based upon presentation given at Naval Postgraduate School).

39. Webster's II New Riverside Dictionary. Revised Ed.(1996) Houghton Mifflin Company. Boston, MA.

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California

3. Dr. Dan C. Boger
   Naval Postgraduate School
   Monterey, California

4. Steven J. Iatrou
   Naval Postgraduate School
   Monterey, California

5. LtCol Karl Pffiefer
   Naval Postgraduate School
   Monterey, CA

6. Director, Information Operation Center For Excellence
   Naval Postgraduate School
   Monterey, California